

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



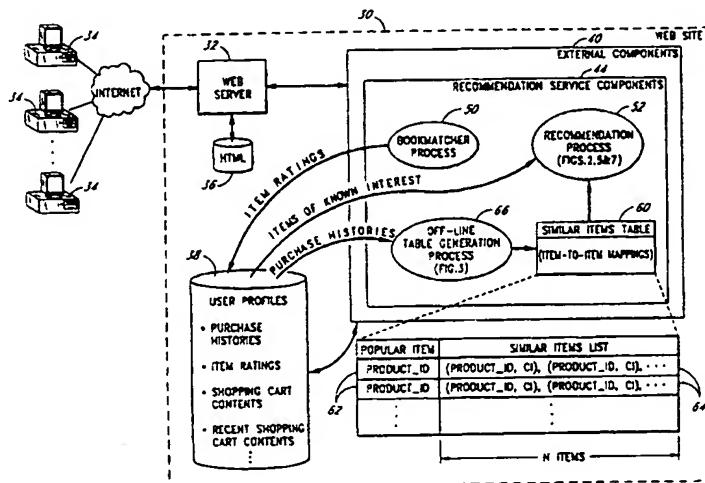
INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ :	A1	(11) International Publication Number:	WO 00/17792
G06F 17/60		(43) International Publication Date:	30 March 2000 (30.03.00)
<p>(21) International Application Number: PCT/US99/20974</p> <p>(22) International Filing Date: 10 September 1999 (10.09.99)</p> <p>(30) Priority Data: 09/157,198 18 September 1998 (18.09.98) US</p> <p>(71) Applicant: AMAZON.COM [US/US]; Suite 1200, 1200 12th Avenue South, Seattle, WA 98144 (US).</p> <p>(72) Inventors: LINDEN, Gregory, D.; 8045 36th Avenue N.E., Seattle, WA 98115 (US). JACOBI, Jennifer, A.; 610 North 65th Street, Seattle, WA 98103 (US). BENSON, Eric, A.; 1953 6th Avenue West, Seattle, WA 98119 (US).</p> <p>(74) Agent: ALTMAN, Daniel, E.; Knobbe, Martens, Olson & Bear, LLP, 16th floor, 620 Newport Center Drive, Newport Beach, CA 92660 (US).</p>		<p>(81) Designated States: AE, AL, AM, AT, AT (Utility model), AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, CZ (Utility model), DE, DE (Utility model), DK, DK (Utility model), DM, EE, EE (Utility model), ES, FI, FI (Utility model), GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SK (Utility model), SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).</p>	

Published

With international search report.

(54) Title: COLLABORATIVE RECOMMENDATIONS USING ITEM-TO-ITEM SIMILARITY MAPPINGS



(57) Abstract

A recommendations service recommends items to individual users based on a set of items that are known to be of interest to the user, such as a set of items previously purchased by the user. The service is used to recommend products to users of a merchant's Web site (30). The service generates the recommendations using a previously-generated table (60) which maps items (62) to lists (64) of "similar" items. The similarities reflected by the table (60) are based on the collective interests of the community of users. To generate personal recommendations, the service retrieves from the table (60) the similar items lists (64) corresponding to the items known to be of interest to the user. These similar items lists (64) are appropriately combined into a single list, which is then sorted and filtered to generate a list of recommended items. Also disclosed are various methods for using the current and/or past contents of a user's electronic shopping cart to generate recommendations.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon	KR	Republic of Korea	PL	Poland		
CN	China	KZ	Kazakhstan	PT	Portugal		
CU	Cuba	LC	Saint Lucia	RO	Romania		
CZ	Czech Republic	LI	Liechtenstein	RU	Russian Federation		
DE	Germany	LK	Sri Lanka	SD	Sudan		
DK	Denmark	LR	Liberia	SE	Sweden		
EE	Estonia			SG	Singapore		

**COLLABORATIVE RECOMMENDATIONS
USING ITEM-TO-ITEM SIMILARITY MAPPINGS**

FIELD OF THE INVENTION

The present invention relates to information filtering and recommendation systems.
5 More specifically, the invention relates to methods for predicting the interests of individual users based on the known interests of a community of users.

BACKGROUND OF THE INVENTION

A recommendation service is a computer-implemented service that recommends items from a database of items. The recommendations are customized to particular users based on 10 information known about the users. One common application for recommendation services involves recommending products to online customers. For example, online merchants commonly provide services for recommending products (books, compact discs, videos, etc.) to customers based on profiles that have been developed for such customers. Recommendation services are also common for recommending Web sites, articles, and other types of 15 informational content to users.

One technique commonly used by recommendation services is known as content-based filtering. Pure content-based systems operate by attempting to identify items which, based on an analysis of item content, are similar to items that are known to be of interest to the user. For example, a content-based Web site recommendation service may operate by parsing 20 the user's favorite Web pages to generate a profile of commonly-occurring terms, and then use this profile to search for other Web pages that include some or all of these terms.

Content-based systems have several significant limitations. For example, content-based methods generally do not provide any mechanism for evaluating the quality or popularity of an item. In addition, content-based methods generally require that the items 25 include some form of content that is amenable to feature extraction algorithms; as a result, content-based systems tend to be poorly suited for recommending movies, music titles, authors, restaurants, and other types of items that have little or no useful, parseable content.

Another common recommendation technique is known as collaborative filtering. In a pure collaborative system, items are recommended to users based on the interests of a 30 community of users, without any analysis of item content. Collaborative systems commonly operate by having the users rate individual items from a list of popular items. Through this process, each user builds a personal profile of ratings data. To generate recommendations for a particular user, the user's profile is initially compared to the profiles of other users to identify one or more "similar users." Items that were rated highly by these similar users (but

which have not yet been rated by the user) are then recommended to the user. An important benefit of collaborative filtering is that it overcomes the above-noted deficiencies of content-based filtering.

As with content-based filtering methods, however, existing collaborative filtering techniques have several problems. One problem is that the user is commonly faced with the onerous task of having to rate items in the database to build up a personal ratings profile. This task can be frustrating, particularly if the user is not familiar with many of the items that are presented for rating purposes. Further, because collaborative filtering relies on the existence of other, similar users, collaborative systems tend to be poorly suited for providing recommendations to users that have unusual tastes.

Another problem with collaborative filtering techniques is that an item in the database normally cannot be recommended until the item has been rated. As a result, the operator of a new collaborative recommendation system is commonly faced with a "cold start" problem in which the service cannot be brought online in a useful form until a threshold quantity of ratings data has been collected. In addition, even after the service has been brought online, it may take months or years before a significant quantity of the database items can be recommended.

Another problem with collaborative filtering methods is that the task of comparing user profiles tends to be time consuming — particularly if the number of users is large (e.g., tens or hundreds of thousands). As a result, a tradeoff tends to exist between response time and breadth of analysis. For example, in a recommendation system that generates real-time recommendations in response to requests from users, it may not be feasible to compare the user's ratings profile to those of all other users. A relatively shallow analysis of the available data (leading to poor recommendations) may therefore be performed.

Another problem with both collaborative and content-based systems is that they generally do not reflect the current preferences of the community of users. In the context of a system that recommends products to customers, for example, there is typically no mechanism for favoring items that are currently "hot sellers." In addition, existing systems do not provide a mechanism for recognizing that the user may be searching for a particular type or category of item.

SUMMARY OF THE DISCLOSURE

The present invention addresses these and other problems by providing a computer-implemented service and associated methods for generating personalized recommendations of items based on the collective interests of a community of users. An important benefit of the

service is that the recommendations are generated without the need for the user, or any other users, to rate items. Another important benefit is that the recommended items are identified using a previously-generated table or other mapping structure which maps individual items to lists of "similar" items. The item similarities reflected by the table are based at least upon 5 correlations between the interests of users in particular items.

The types of items that can be recommended by the service include, without limitation, books, compact discs ("CDs"), videos, authors, artists, item categories, Web sites, and chat groups. The service may be implemented, for example, as part of a Web site, online services network, e-mail notification service, document filtering system, or other type of computer 10 system that explicitly or implicitly recommends items to users. In a preferred embodiment described herein, the service is used to recommend works such as book titles and music titles to users of an online merchant's Web site.

In accordance with one aspect of the invention, the mappings of items to similar items ("item-to-item mappings") are generated periodically, such as once per week, by an off-line 15 process which identifies correlations between known interests of users in particular items. For example, in the embodiment described in detail below, the mappings are generated by periodically analyzing user purchase histories to identify correlations between purchases of items. The similarity between two items is preferably measured by determining the number of users that have an interest in both items relative to the number of users that have an interest in 20 either item (e.g., items A and B are highly similar because a relatively large portion of the users that bought one of the items also bought the other item). The item-to-item mappings could also incorporate other types of similarities, including content-based similarities extracted by analyzing item descriptions or content.

To generate a set of recommendations for a given user, the service retrieves from the 25 table the similar items lists corresponding to items already known to be of interest to the user, and then appropriately combines these lists to generate a list of recommended items. For example, if there are three items that are known to be of interest to the user (such as three items the user recently purchased), the service may retrieve the similar items lists for these three items from the table and combine these lists. Because the item-to-item mappings are 30 regenerated periodically based on up-to-date sales data, the recommendations tend to reflect the current buying trends of the community.

In accordance with another aspect of the invention, the similar items lists read from the table may be appropriately weighted (prior to being combined) based on indicia of the user's affinity for, or current interest in, the corresponding items of known interest. For example, the

similar items list for a book that was purchased in the last week may be weighted more heavily than the similar items list for a book that was purchased four months ago. Weighting a similar items list heavily has the effect of increasing the likelihood that the items in that list will be included in the recommendations that are ultimately presented to the user.

5 An important aspect of the service is that the relatively computation-intensive task of correlating item interests is performed off-line, and the results of this task (item-to-item mappings) stored in a mapping structure for subsequent look-up. This enables the personal recommendations to be generated rapidly and efficiently (such as in real-time in response to a request by the user), without sacrificing breadth of analysis.

10 Another feature of the invention involves using the current and/or recent contents of the user's shopping cart as inputs to the recommendation service (or to another type of recommendation service which generates recommendations given a unary listing of items). For example, if the user currently has three items in his or her shopping cart, these three items can be treated as the items of known interest for purposes of generating recommendations, in
15 which case the recommendations may be generated and displayed automatically when the user views the shopping cart contents. Using the current and/or recent shopping cart contents as inputs tends to produce recommendations that are highly correlated to the current short-term interests of the user — even if these short term interest differ significantly from the user's general preferences. For example, if the user is currently searching for books on a particular
20 topic and has added several such books to the shopping cart, this method will more likely produce other books that involve the same or similar topics.

Another feature of the invention involves allowing the user to create multiple shopping carts under a single account (such as shopping carts for different family members), and generating recommendations that are specific to a particular shopping cart. For example, the
25 user can be prompted to select a particular shopping cart (or set of shopping carts), and the recommendations can then be generated based on the items that were purchased from or otherwise placed into the designated shopping cart(s). This feature of the invention allows users to obtain recommendations that correspond to the role or purpose (e.g., work versus pleasure) of a particular shopping cart.

30 Two specific implementations of the service are disclosed, both of which generate personal recommendations using the same type of table. In the first implementation, the recommendations are based on the items that have recently been rated or purchased by the user. In the second implementation, the recommendations are based on the current shopping cart contents of the user.

BRIEF DESCRIPTION OF THE DRAWINGS

These and other features of the invention will now be described with reference to the drawings summarized below. These drawings and the associated description are provided to illustrate a preferred embodiment of the invention, and not to limit the scope of the invention.

5 Figure 1 illustrates a Web site which implements a recommendation service which operates in accordance with the invention, and illustrates the flow of information between components.

Figure 2 illustrates a sequence of steps that are performed by the recommendation process of Figure 1 to generate personalized recommendations.

10 Figure 3 illustrates a sequence of steps that are performed by the table generation process of Figure 1 to generate a similar items table, and illustrates temporary data structures generated during the process.

Figure 4 is a Venn diagram illustrating a hypothetical purchase history profile of three items.

15 Figure 5 illustrates one specific implementation of the sequence of steps of Figure 2.

Figure 6 illustrates the general form of a Web pages used to present the recommendations of the Figure 5 process to the user.

Figure 7 illustrates another specific implementation of the sequence of steps of Figure 2.

20 DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

The various features and methods of the invention will now be described in the context of a recommendation service, including two specific implementations thereof, that is used to recommend book titles, music titles, video titles, and other types of items to individual users of the Amazon.com Web site. As will be recognized to those skilled in the art, the disclosed methods can also be used to recommend other types of items, including non-physical items. By way of example and not limitation, the disclosed methods can also be used to recommend authors, artists, categories or groups of titles, Web sites, chat groups, movies, television shows, downloadable content, restaurants, and other users.

Throughout the description, reference will be made to various implementation-specific details of the recommendation service, the Amazon.com Web site, and other recommendation services of the Web site. These details are provided in order to fully illustrate preferred embodiments of the invention, and not to limit the scope of the invention. The scope of the invention is set forth in the appended claims.

I. Overview of Web Site and Recommendation Services

The Amazon.com Web site includes functionality for allowing users to search, browse, and make purchases from an online catalog of several million book titles, music titles, video titles, and other types of items. Using a shopping cart feature of the site, users can add and remove items to/from a personal shopping cart which is persistent over multiple sessions. (As used herein, a "shopping cart" is a data structure and associated code which keeps track of items that have been selected by a user for possible purchase.) For example, a user can modify the contents of the shopping cart over a period of time, such as one week, and then proceed to a check out area of the site to purchase the shopping cart contents.

10 The user can also create multiple shopping carts within a single account. For example, a user can set up separate shopping carts for work and home, or can set up separate shopping carts for each member of the user's family. A preferred shopping cart scheme for allowing users to set up and use multiple shopping carts is disclosed in U.S. Appl. No. 09/104,942, filed June 25, 1998, titled METHOD AND SYSTEM FOR ELECTRONIC COMMERCE USING
15 MULTIPLE ROLES.

The site also implements a variety of different recommendation services for recommending book titles, music titles, and/or video titles to users. One such service, known as BookMatcher™, allows users to interactively rate individual books on a scale of 1-5 to create personal item ratings profiles, and applies collaborative filtering techniques to these
20 profiles to generate personal recommendations. The BookMatcher service is described in detail in U.S. Appl. No. 09/040,171 filed March 17, 1998. The site may also include associated services that allow users to rate other types of items, such as CDs and videos. As described below, the ratings data collected by the BookMatcher service and similar services is optionally incorporated into the recommendation processes of the present invention.

25 Another type of service is a recommendation service which operates in accordance with the invention. The service ("Recommendation Service") is preferably used to recommend book titles, music titles and/or video titles to users, but could also be used in the context of the same Web site to recommend other types of items, including authors, artists, and groups or categories of titles. Briefly, given a unary listing of items that are "known" to
30 be of interest to a user (e.g., a list of items purchased, rated, and/or viewed by the user), the Recommendation Service generates a list of additional items ("recommendations") that are predicted to be of interest to the user. (As used herein, the term "interest" refers generally to a user's liking of or affinity for an item; the term "known" is used to distinguish items for which

the user has implicitly or explicitly indicated some level of interest from items predicted by the Recommendation Service to be of interest.)

The recommendations are generated using a table which maps items to lists of “similar” items (“similar items lists”), without the need for users to rate any items (although 5 ratings data may optionally be used). For example, if there are three items that are known to be of interest to a particular user (such as three items the user recently purchased), the service may retrieve the similar items lists for these three items from the table, and appropriately combine these lists (as described below) to generate the recommendations.

In accordance with one aspect of the invention, the mappings of items to similar items 10 (“item-to-item mappings”) are generated periodically, such as once per week, from data which reflects the collective interests of the community of users. More specifically, the item-to-item mappings are generated by an off-line process which identifies correlations between known interests of users in particular items. For example, in the embodiment described in detail below, the mappings are generated by analyzing user purchase histories to identify 15 correlations between purchases of particular items (e.g., items A and B are similar because a relatively large portion of the users that purchased item A also bought item B). The item-to-item mappings could also reflect other types of similarities, including content-based similarities extracted by analyzing item descriptions or content.

An important aspect of the Recommendation Service is that the relatively 20 computation-intensive task of correlating item interests is performed off-line, and the results of this task (item-to-item mappings) are stored in a mapping structure for subsequent look-up. This enables the personal recommendations to be generated rapidly and efficiently (such as in real-time in response to a request by the user), without sacrificing breadth of analysis.

In accordance with another aspect of the invention, the similar items lists read from the 25 table are appropriately weighted (prior to being combined) based on indicia of the user’s affinity for or current interest in the corresponding items of known interest. For example, in one embodiment described below, if the item of known interest was previously rated by the user (such as through use of the BookMatcher service), the rating is used to weight the corresponding similar items list. Similarly, the similar items list for a book that was purchased 30 in the last week may be weighted more heavily than the similar items list for a book that was purchased four months ago.

Another feature of the invention involves using the current and/or recent contents of the user’s shopping cart as inputs to the Recommendation Service. For example, if the user currently has three items in his or her shopping cart, these three items can be treated as the

items of known interest for purposes of generating recommendations, in which case the recommendations may be generated and displayed automatically when the user views the shopping cart contents. If the user has multiple shopping carts, the recommendations are preferably generated based on the contents of the shopping cart implicitly or explicitly 5 designated by the user, such as the shopping cart currently being viewed. This method of generating recommendations can also be used within other types of recommendation systems, including content-based systems and systems that do not use item-to-item mappings.

Using the current and/or recent shopping cart contents as inputs tends to produce recommendations that are highly correlated to the current short-term interests of the user — 10 even if these short term interests are not reflected by the user's purchase history. For example, if the user is currently searching for a father's day gift and has selected several books for prospective purchase, this method will have a tendency to identify other books that are well suited for the gift recipient.

Another feature of the invention involves generating recommendations that are 15 specific to a particular shopping cart. This allows a user who has created multiple shopping carts to conveniently obtain recommendations that are specific to the role or purpose to the particular cart. For example, a user who has created a personal shopping cart for buying books for her children can designate this shopping cart to obtain recommendations of children's books. In one embodiment of this feature, the recommendations are generated 20 based solely upon the current contents of the shopping cart selected for display. In another embodiment, the user may designate one or more shopping carts to be used to generate the recommendations, and the service then uses the items that were purchased from these shopping carts as the items of known interest.

As will be recognized by those skilled in the art, the above-described techniques for 25 using shopping cart contents to generate recommendations can also be incorporated into other types of recommendation systems, including pure content-based systems.

Figure 1 illustrates the basic components of the Amazon.com Web site 30, including the components used to implement the Recommendation Service. The arrows in Figure 1 show the general flow of information that is used by the Recommendation Service. As 30 illustrated by Figure 1, the Web site 30 includes a Web server application 32 ("Web server") which processes HTTP (Hypertext Transfer Protocol) requests received over the Internet from user computers 34. The Web server 34 accesses a database 36 of HTML (Hypertext Markup Language) content which includes product information pages and other browsable information about the various products of the catalog. The "items" that are the subject of the

Recommendation Service are the titles (regardless of media format such as hardcover or paperback) that are represented within this database 36.

The Web site 30 also includes a "user profiles" database 38 which stores account-specific information about users of the site. Because a group of individuals can share an account, a given "user" from the perspective of the Web site may include multiple actual users. As illustrated by Figure 1, the data stored for each user may include one or more of the following types of information (among other things) that can be used to generate recommendations in accordance with the invention: (a) the user's purchase history, including dates of purchase, (b) the user's item ratings profile (if any), (c) the current contents of the user's personal shopping cart(s), and (d) a listing of items that were recently (e.g., within the last six months) removed from the shopping cart(s) without being purchased ("recent shopping cart contents"). If a given user has multiple shopping carts, the purchase history for that user may include information about the particular shopping cart used to make each purchase; preserving such information allows the Recommendation Service to be configured to generate recommendations that are specific to a particular shopping cart.

As depicted by Figure 1, the Web server 32 communicates with various external components 40 of the site. These external components 40 include, for example, a search engine and associated database (not shown) for enabling users to interactively search the catalog for particular items. Also included within the external components 40 are various order processing modules (not shown) for accepting and processing orders, and for updating the purchase histories of the users.

The external components 40 also include a shopping cart process (not shown) which adds and removes items from the users' personal shopping carts based on the actions of the respective users. (The term "process" is used herein to refer generally to one or more code modules that are executed by a computer system to perform a particular task or set of related tasks.) In one embodiment, the shopping cart process periodically "prunes" the personal shopping cart listings of items that are deemed to be dormant, such as items that have not been purchased or viewed by the particular user for a predetermined period of time (e.g. two weeks). The shopping cart process also preferably generates and maintains the user-specific listings of recent shopping cart contents.

The external components 40 also include recommendation service components 44 that are used to implement the site's various recommendation services. Recommendations generated by the recommendation services are returned to the Web server 32, which incorporates the recommendations into personalized Web pages transmitted to users.

The recommendation service components 44 include a BookMatcher application 50 which implements the above-described BookMatcher service. Users of the BookMatcher service are provided the opportunity to rate individual book titles from a list of popular titles. The book titles are rated according to the following scale:

5

- 1 = Bad!
- 2 = Not for me
- 3 = OK
- 4 = Liked it
- 5 = Loved it!

10

Users can also rate book titles during ordinary browsing of the site. As depicted in Figure 1, the BookMatcher application 50 records the ratings within the user's items rating profile. For example, if a user of the BookMatcher service gives the book *Into Thin Air* a score of "5," the 15 BookMatcher application 50 would record the item (by ISBN or other identifier) and the score within the user's item ratings profile. The BookMatcher application 50 uses the users' item ratings profiles to generate personal recommendations, which can be requested by the user by selecting an appropriate hyperlink. As described in detail below, the item ratings profiles are also used by an "Instant Recommendations" implementation of the Recommendation Service.

20

The recommendation services components 44 also include a recommendation process 52, a similar items table 60, and an off-line table generation process 66, which collectively implement the Recommendation Service. As depicted by the arrows in Figure 1, the recommendation process 52 generates personal recommendations based on information stored within the similar items table 60, and based on the items that are known to be of interest 25 ("items of known interest") to the particular user.

25

In the embodiments described in detail below, the items of known interest are identified based on information stored in the user's profile, such as by selecting all items purchased by the user or all items in the user's shopping cart. In other embodiments of the invention, other types of methods or sources of information could be used to identify the items 30 of known interest. For example, in a service used to recommend Web sites, the items (Web sites) known to be of interest to a user could be identified by parsing a Web server access log and/or by extracting URLs from the "favorite places" list of the user's Web browser. In a service used to recommend restaurants, the items (restaurants) of known interest could be identified by parsing the user's credit card records to identify restaurants that were visited 35 more than once.

The various processes 50, 52, 66 of the recommendation services may run, for example, on one or more Unix or NT based workstations or physical servers (not shown) of the Web site 30. The similar items table 60 is preferably stored as a B-tree data structure to permit efficient look-up, and may be replicated across multiple machines (together with the 5 associated code of the recommendation process 52) to accommodate heavy loads.

II. Similar Items Table (Figure 1)

The general form and content of the similar items table 60 will now be described with reference to Figure 1. As this table can take on many alternative forms, the details of the table are intended to illustrate, and not limit, the scope of the invention.

10 As indicated above, the similar items table 60 maps items to lists of similar items based at least upon the collective interests of the community of users. The similar items table 60 is preferably generated periodically (e.g., once per week) by the off-line table generation process 66. The table generation process 66 generates the table 60 from data that reflects the collective interests of the community of users. In the embodiment described in detail herein, 15 the similar items table is generated exclusively from the purchase histories of the community of users (as depicted in Figure 1). In other embodiments, the table 60 may additionally or alternatively be generated from other indicia of user-item interests, including indica based on users viewing activities, shopping cart activities, and item rating profiles. For example, the table 60 could be built exclusively from the present and/or recent shopping cart contents of 20 users. The similar items table 60 could also reflect non-collaborative type item similarities, including content-based similarities derived by comparing item contents or descriptions.

Each entry in the similar items table 60 is preferably in the form of a mapping of a popular item 62 to a corresponding list 64 of similar items ("similar items lists"). As used herein, a "popular" item is an item which satisfies some pre-specified popularity criteria. For 25 example, in the embodiment described herein, an item is treated as popular if it has been purchased by more than 30 customers during the life of the Web site. Using this criteria produces a set of popular items (and thus a recommendation service) which grows over time. The similar items list 64 for a given popular item 62 may include other popular items.

In other embodiments involving sales of products, the table 60 may include entries for 30 most or all of the products of the online merchant, rather than just the popular items. In the embodiment described herein, several different types of items (books, CDs, videos, etc.) are reflected within the same table 60, although separate tables could alternatively be generated for each type of item.

Each similar items list 64 consists of the N (e.g., 20) items which, based on correlations between purchases of items, are deemed to be the most closely related to the respective popular item 62. Each item in the similar items list 64 is stored together with a commonality index ("CI") value which indicates the relatedness of that item to the popular 5 item 62, based on sales of the respective items. A relatively high commonality index for a pair of items ITEM A and ITEM B indicates that a relatively large percentage of users who bought ITEM A also bought ITEM B (and vice versa). A relatively low commonality index for ITEM A and ITEM B indicates that a relatively small percentage of the users who bought ITEM A also bought ITEM B (and vice versa). As described below, the similar items lists are 10 generated, for each popular item, by selecting the N other items that have the highest commonality index values. Using this method, ITEM A may be included in ITEM B's similar items list even though ITEM B is not present in ITEM A's similar items list.

In the embodiment depicted by Figure 1, the items are represented within the similar items table 60 using product IDs, such as ISBNs or other identifiers. Alternatively, the items 15 could be represented within the table by title ID, where each title ID corresponds to a given "work" regardless of its media format. In either case, different items which correspond to the same work, such as the hardcover and paperback versions of a given book or the VCR cassette and DVD versions of a given video, are preferably treated as a unit for purposes of generating recommendations.

20 Although the recommendable items in the described system are in the form of book titles, music titles and videos titles, it will be appreciated that the underlying methods and data structures can be used to recommend a wide range of other types of items. For example, in the system depicted by Figure 1, the Recommendation Service could also be used to recommend authors, artists, and categorizations or groups of works.

25 III. General Process for Generating Recommendations (Figure 2)

The general sequence of steps that are performed by the recommendation process 52 to generate a set of personal recommendations will now be described with reference to Figure 2. This process, and the more specific implementations of the process depicted by Figures 5 and 7 (described below), are intended to illustrate, and not limit, the scope of the invention.

30 The Figure 2 process is preferably invoked in real-time in response to an online action of the user. For example, in an Instant Recommendations implementation (Figure 5 and 6) of the service, the recommendations are generated and displayed in real-time (based on the user's purchase history and/or item ratings profile) in response to selection by the user of a corresponding hyperlink, such as a hyperlink which reads "Instant Book Recommendations"

or "Instant Music Recommendations." In a shopping cart based implementation (Figure 7), the recommendations are generated (based on the user's current and/or recent shopping cart contents) in real-time when the user initiates a display of a shopping cart, and are displayed on the same Web page as the shopping cart contents. The Instant Recommendations and 5 shopping cart based embodiments are described separately below under corresponding headings.

Any of a variety of other methods can be used to initiate the recommendations generation process and to display the recommendations to the user. For example, the recommendations can automatically be generated periodically and sent to the user by e-mail, 10 in which case the e-mail listing may contain hyperlinks to the product information pages of the recommended items. Further, the personal recommendations could be generated in advance of any request or action by the user, and cached by the Web site 30 until requested.

As illustrated by Figure 2, the first step (step 80) of the recommendations-generation process involves identifying a set of items that are of known interest to the user. The 15 "knowledge" of the user's interest can be based on explicit indications of interest (e.g., the user rated the item highly) or implicit indications of interest (e.g., the user added the item to a shopping cart). Items that are not "popular items" within the similar items table 60 can optionally be ignored during this step.

In the embodiment depicted in Figure 1, the items of known interest are selected from 20 one or more of the following groups: (a) items in the user's purchase history (optionally limited to those items purchased from a particular shopping cart); (b) items in the user's shopping cart (or a particular shopping cart designated by the user), (c) items rated by the user (optionally with a score that exceeds a certain threshold, such as two), and (d) items in the "recent shopping cart contents" list associated with a given user or shopping cart. In other 25 embodiments, the items of known interest may additionally or alternatively be selected based on the viewing activities of the user. For example, the recommendations process 52 could select items that were viewed by the user for an extended period of time and/or viewed more than once. Further, the user could be prompted to select items of interest from a list of popular items.

30 For each item of known interest, the service retrieves the corresponding similar items list 64 from the similar items table 60 (step 82), if such a list exists. If no entries exist in the table 60 for any of the items of known interest, the process 52 may be terminated; alternatively, the process could attempt to identify additional items of interest, such as by accessing other sources of interest information.

In step 84, the similar items lists 64 are optionally weighted based on information about the user's affinity for the corresponding items of known interest. For example, a similar items list 64 may be weighted heavily if the user gave the corresponding popular item a rating of "5" on a scale of 1-5, or if the user purchased multiple copies of the item. Weighting a similar items list 64 heavily has the effect of increasing the likelihood that the items in that list will be included in the recommendations that are ultimately presented to the user. In one implementation described below, the user is presumed to have a greater affinity for recently purchased items over earlier purchased items.

The similar items lists 64 are preferably weighted by multiplying the commonality index values of the list by a weighting value. The commonality index values as weighted by any applicable weighting value are referred to herein as "scores." In other embodiments, the recommendations may be generated without weighting the similar items lists 64.

If multiple similar items lists 64 are retrieved in step 82, the lists are appropriately combined (step 86), such as by merging the lists while summing the scores of like items. The resulting list is then sorted (step 88) in order of highest-to-lowest score. In step 90, the sorted list is filtered to remove unwanted items. The items removed during the filtering process may include, for example, items that have already been purchased or rated by the user, and items that fall outside any product group (such as music or books), product category (such as non-fiction), or content rating (such as PG or adult) designated by the user. The filtering step could alternatively be performed at a different stage of the process, such as during the retrieval of the similar items lists from the table 60. The result of step 90 is a list ("recommendations list") of other items to be recommended to the user.

In step 92, one or more additional items are optionally added to the recommendations list. In one embodiment, the items added in step 92 are selected from the set of items (if any) in the user's "recent shopping cart contents" list. As an important benefit of this step, the recommendations include one or more items that the user previously considered purchasing but did not purchase. The items added in step 92 may additionally or alternatively be selected using another recommendations method, such as a content-based method.

Finally, in step 94, a list of the top M (e.g., 15) items of the recommendations list are returned to the Web server 32 (Figure 1). The Web server incorporates this list into one or more Web pages that are returned to the user, with each recommended item being presented as a hypertextual link to the item's product information page. The recommendations may alternatively be conveyed to the user by email, facsimile, or other transmission method.

Further, the recommendations could be presented as advertisements for the recommended items.

IV. Generation of Similar Items Table (Figures 3 and 4)

The table-generation process 66 is preferably executed periodically (e.g., once a week) 5 to generate a similar items table 60 that reflects the most recent purchase history data. The recommendation process 52 uses the most recently generated version of the table 60 to generate recommendations.

Figure 3 illustrates the sequence of steps that are performed by the table generation process 66 to build the similar items table 60. The general form of temporary data structures 10 that are generated during the process are shown at the right of the drawing. As will be appreciated by those skilled in the art, any of a variety of alternative methods could be used to generate the table 60.

As depicted by Figure 3, the process initially retrieves the purchase histories for all customers (step 100). Each purchase history is in the general form of the user ID of a 15 customer together with a list of the product IDs (ISBNs, etc.) of the items (books, CDs, videos, etc.) purchased by that customer. In embodiments which support multiple shopping carts within a given account, each shopping cart could be treated as a separate customer for purposes of generating the table. For example, if a given user (or group of users that share an account) purchased items from two different shopping carts within the same account, these 20 purchases could be treated as the purchases of separate users.

The product IDs may be converted to title IDs during this process, or when the table 60 is later used to generate recommendations, so that different versions of an item (e.g., hardcover and paperback) are represented as a single item. This may be accomplished, for example, by using a separate database which maps product IDs to title IDs. To generate a 25 similar items table that strongly reflects the current tastes of the community, the purchase histories retrieved in step 100 can be limited to a specific time period, such as the last six months.

In steps 102 and 104, the process generates two temporary tables 102A and 104A. The first table 102A maps individual customers to the items they purchased. The second table 30 104A maps items to the customers that purchased such items. To avoid the effects of "ballot stuffing," multiple copies of the same item purchased by a single customer are represented with a single table entry. For example, even if a single customer purchased 4000 copies of one book, the customer will be treated as having purchased only a single copy. In addition,

items that were sold to an insignificant number (e.g., < 15) of customers are preferably omitted or deleted from the tables 102A, 104B.

In step 106, the process identifies the items that constitute "popular" items. This may be accomplished, for example, by selecting from the item-to-customers table 104A those 5 items that were purchased by more than a threshold number (e.g., 30) of customers. In the context of the Amazon.com Web site, to resulting set of popular items may contain hundreds of thousands or millions of items.

In step 108, the process counts, for each (popular_item, other_item) pair, the number of customers that are in common. A pseudocode sequence for performing this step is listed in 10 Table 1. The result of step 108 is a table that indicates, for each (popular_item, other_item) pair, the number of customers the two have in common. For example, in the hypothetical table 108A of Figure 3, POPULAR_A and ITEM_B have seventy customers in common, indicating that seventy customers bought both items.

TABLE 1

<pre> for each popular_item for each customer in customers of item for each other_item in items of customer increment common-customer-count(popular_item, other_item) </pre>
--

15

In step 110, the process generates the commonality indexes for each (popular_item, other_item) pair in the table 108A. As indicated above, the commonality index (CI) values are measures of the similarity between two items, with larger CI values indicating greater degrees of similarity. The commonality indexes are preferably generated such that, for a 20 given popular_item, the respective commonality indexes of the corresponding other_items take into consideration both (a) the number of customers that are common to both items, and (b) the total number of customers of the other_item. A preferred method for generating the commonality index values is set forth in the equation below.

$$25 \quad CI(item_A, item_B) = \frac{\text{customers of item_A and item_B}}{\sqrt{(\text{customers of item_A}) \times (\text{customers of item_B})}}$$

Figure 4 illustrates this method in example form. In the Figure 4 example, item_P (a popular item) has two "other items," item_X and item_Y. Item_P has been purchased by 300

customers, item_X by 300 customers, and item_Y by 30,000 customers. In addition, item_P and item_X have 20 customers in common, and item_P and item_Y have 25 customers in common. Applying the equation above to the values shown in Figure 4 produces the following results:

5

$$CI(item_P, item_X) = 20/\sqrt{300 \times 300}) = 0.0667$$

$$CI(item_P, item_Y) = 25/\sqrt{300 \times 30,000}) = 0.0083$$

Thus, even though items P and Y have more customers in common than items P and X, items 10 P and X are treated as being more similar than items P and Y. This result desirably reflects the fact that the percentage of item_X customers that bought item_P (6.7%) is much greater than the percentage of item_Y customers that bought item_P (0.08%).

Because this equation is symmetrical (i.e., $CI(item_A, item_B) = CI(item_B, item_A)$), it is not necessary to separately calculate the CI value for every location in the table 15 108A. In other embodiments, an asymmetrical method may be used to generate the CI values. For example, the CI value for a (popular_item, other_item) pair could be generated as (customers of popular_item and other_item)/(customers of other_item).

Following step 110 of Figure 3, each popular item has a respective "other_items" list which includes all of the other_items from the table 108A and their associated CI values. In 20 step 112, each other_items list is sorted from highest-to-lowest commonality index. Using the Figure 4 values as an example, item_X would be positioned closer to the top of the item_B's list than item_Y, since $0.014907 > 0.001643$.

In step 114, the sorted other_items lists are filtered by deleting all list entries that have fewer than 3 customers in common. For example, in the other_items list for POPULAR_A in 25 table 108A, ITEM_A would be deleted since POPULAR_A and ITEM_A have only two customers in common. Deleting such entries tends to reduce statistically poor correlations between item sales.

In step 116, the sorted other_items lists are truncated to length N to generate the similar items lists, and the similar items lists are stored in a B-tree table structure for efficient 30 look-up.

As indicated above, any of a variety of other methods for evaluating similarities between items could be incorporated into the table generation process 66. For example, the table generation process could compare item contents and/or use previously-assigned product categorizations as additional indicators of item similarities. An important benefit of the Figure

3 method, however, is that the items need not contain any content that is amenable to feature extraction techniques, and need not be pre-assigned to any categories. For example, the method can be used to generate a similar items table given nothing more than the product IDs of a set of products and user purchase histories with respect to these products.

5 Another important benefit of the Recommendation Service is that the bulk of the processing (the generation of the similar items table 60) is performed by an off-line process. Once this table has been generated, personalized recommendations can be generated rapidly and efficiently, without sacrificing breadth of analysis.

V. Instant Recommendations Service (Figures 5 and 6)

10 A specific implementation of the Recommendation Service, referred to herein as the Instant Recommendations service, will now be described with reference to Figures 5 and 6.

As indicated above, the Instant Recommendations service is invoked by the user by selecting a corresponding hyperlink from a Web page. For example, the user may select an "Instant Book Recommendations" or similar hyperlink to obtain a listing of recommended
15 book titles, or may select a "Instant Music Recommendations" or "Instant Video Recommendations" hyperlink to obtain a listing of recommended music or video titles. As described below, the user can also request that the recommendations be limited to a particular item category, such as "non-fiction," "jazz" or "comedies." The Instant Recommendations service generates the recommendations based exclusively on the purchase history and any
20 item ratings profile of the particular user. The service becomes available to the user (i.e., the appropriate hyperlink is presented to the user) once the user has purchased and/or rated a threshold number (e.g. three) of popular items within the corresponding product group. If the user has established multiple shopping carts, the user may also be presented the option of designating a particular shopping cart to be used in generating the recommendations.

25 Figure 5 illustrates the sequence of steps that are performed by the Instant Recommendations service to generate personal recommendations. Steps 180-194 in Figure 5 correspond, respectively, to steps 80-94 in Figure 2. In step 180, the process 52 identifies all popular items that have been purchased by the user (from a particular shopping cart, if designated) or rated by the user, within the last six months. In step 182, the process retrieves
30 the similar items lists 64 for these popular items from the similar items table 60.

In step 184, the process 52 weights each similar items list based on the duration since the associated popular item was purchased by the user (with recently-purchased items weighted more heavily), or if the popular item was not purchased, the rating given to the popular item by the user. The formula used to generate the weight values to apply to each

similar items list is listed in C in Table 2. In this formula, "is_purchased" is a boolean variable which indicates whether the popular item was purchased, "rating" is the rating value (1-5), if any, assigned to the popular item by the user, "order_date" is the date/time (measured in seconds since 1970) the popular item was purchased, "now" is the current date/time
5 (measured in seconds since 1970), and "6 months" is six months in seconds.

TABLE 2

1	$\text{Weight} = (\text{is_purchased} ? 5 : \text{rating}) * 2 - 5 *$
2	$(1 + (\max(\text{is_purchased} ? \text{order_date} : 0) - (\text{now} - 6 \text{ months}), 0))$
3	$/ (6 \text{ months}))$

- In line 1 of the formula, if the popular item was purchased, the value "5" (the maximum possible rating value) is selected; otherwise, the user's rating of the item is selected.
10 The selected value (which may range from 1-5) is then multiplied by 2, and 5 is subtracted from the result. The value calculated in line 1 thus ranges from a minimum of -3 (if the item was rated a "1") to a maximum of 5 (if the item was purchased or was rated a "5").
The value calculated in line 1 is multiplied by the value calculated in lines 2 and 3, which can range from a minimum of 1 (if the item was either not purchased or was purchased
15 at least six months ago) to a maximum of 2 (if order_date = now). Thus, the weight can range from a minimum of -6 to a maximum of 10. Weights of zero and below indicate that the user rated the item a "2" or below. Weights higher than 5 indicate that the user actually purchased the item (although a weight of 5 or less is possible even if the item was purchased), with higher values indicating more recent purchases.
20 The similar items lists 64 are weighted in step 184 by multiplying the CI values of the list by the corresponding weight value. For example, if the weight value for a given popular item is ten, and the similar items list 64 for the popular item is

(productid_A, 0.10), (productid_B, 0.09), (productid_C, 0.08), ...

- 25 the weighted similar items list would be:

(productid_A, 1.0), (productid_B, 0.9), (productid_C, 0.8), ...

- 30 The numerical values in the weighted similar items lists are referred to as "scores."

In step 186, the weighted similar items lists are merged (if multiple lists exist) to form a single list. During this step, the scores of like items are summed. For example, if a given other_item appears in three different similar items lists 64, the three scores (including any negative scores) are summed to produce a composite score.

- 5 In step 188, the resulting list is sorted from highest-to-lowest score. The effect of the sorting operation is to place the most relevant items at the top of the list. In step 190, the list is filtered by deleting any items that (1) have already been purchased or rated by the user, (2) have a negative score, or (3) do not fall within the designated product group (e.g., books) or category (e.g., "science fiction," or "jazz").
- 10 In step 192 one or more items are optionally selected from the recent shopping cart contents list (if such a list exists) for the user, excluding items that have been rated by the user or which fall outside the designated product group or category. The selected items, if any, are inserted at randomly-selected locations within the top M (e.g., 15) positions in the recommendations list. Finally, in step 194, the top M items from the recommendations list are
15 returned to the Web server 32, which incorporates these recommendations into one or more Web pages.

The general form of such a Web page is shown in Figure 6, which lists five recommended items. From this page, the user can select a link associated with one of the recommended items to view the product information page for that item. In addition, the user
20 can select a "more recommendations" button 200 to view additional items from the list of M items. Further, the user can select a "refine your recommendations" link to rate or indicate ownership of the recommended items. Indicating ownership of an item causes the item to be added to the user's purchase history listing.

The user can also select a specific category such as "non-fiction" or "romance" from a
25 drop-down menu 202 to request category-specific recommendations. Designating a specific category causes items in all other categories to be filtered out in step 190 (Figure 5).

VI. Shopping Cart Based Recommendations (Figure 7)

Another specific implementation of the Recommendation Service, referred to herein as shopping cart recommendations, will now be described with reference to Figure 7.

30 The shopping cart recommendations service is preferably invoked automatically when the user displays the contents of a shopping cart that contains more than a threshold number (e.g., 1) of popular items. The service generates the recommendations based exclusively on the current contents of the shopping cart. As a result, the recommendations tend to be highly correlated to the user's current shopping interests. In other implementations, the

recommendations may also be based on other items that are deemed to be of current interest to the user, such as items in the recent shopping cart contents of the user and/or items recently viewed by the user. Further, other indications of the user's current shopping interests could be incorporated into the process. For example, any search terms typed into the site's search engine during the user's browsing session could be captured and used to perform content-based filtering of the recommended items list.

Figure 7 illustrates the sequence of steps that are performed by the shopping cart recommendations service to generate a set of shopping-cart-based recommendations. In step 282, the similar items list for each popular item in the shopping cart is retrieved from the similar items table 60. The similar items list for one or more additional items that are deemed to be of current interest could also be retrieved during this step, such as the list for an item recently deleted from the shopping cart or recently viewed for an extended period of time.

In step 286, these similar items lists are merged while summing the commonality index (CI) values of the list items. In step 288, the resulting list is sorted from highest-to-lowest score. In step 290, the list is filtered to remove any items that exist in the shopping cart or have been purchased or rated by the user. Finally, in step 294, the top M (e.g., 5) items of the list are returned as recommendations. The recommendations are preferably presented to the user on the same Web page (not shown) as the shopping cart contents.

If the user has defined multiple shopping carts, the recommendations generated by the Figure 7 process may be based solely on the contents of the shopping cart currently selected for display. As described above, this allows the user to obtain recommendations that correspond to the role or purpose of a particular shopping cart (e.g., work versus home).

The various uses of shopping cart contents to generate recommendations as described above can be applied to other types of recommendation systems, including content-based systems. For example, the current and/or past contents of a shopping cart can be used to generate recommendations in a system in which mappings of items to lists of similar items are generated from a computer-based comparison of item contents. Methods for performing content-based similarity analyses of items are well known in the art, and are therefore not described herein.

Although this invention has been described in terms of certain preferred embodiments, other embodiments that are apparent to those of ordinary skill in the art are also within the scope of this invention. For example, although the embodiments described herein employ item lists, other programming methods for keeping track of and combining sets of similar

items can be used. Accordingly, the scope of the present invention is intended to be defined only by reference to the appended claims.

In the claims which follow, reference characters used to denote process steps are provided for convenience of description only, and not to imply a particular order for 5 performing the steps.

WHAT IS CLAIMED IS:

1. In a computer system that provides user access to a database of items, a method of predicting items that will be of interest to a user, the method comprising the
5 computer-implemented steps of:
 - (a) generating a data structure which maps individual items of the database to corresponding sets of similar items in which similarities between items are based at least upon the collective item interests of a community of users;
 - (b) identifying items that are known to be of interest to the user;
 - 10 (c) for each of a plurality of the items identified in step (b), accessing the data structure to identify a corresponding set of similar items;
 - (d) combining the sets of similar items identified in step (c) to generate a set of additional items; and
 - (e) selecting at least some of the additional items of the set generated in
15 step (d) to recommend to the user.
2. The method of Claim 1, wherein step (a) comprises analyzing purchase histories of users to identify correlations between purchases of items.
3. The method of Claim 2, wherein step (a) comprises, for each of a plurality of popular items, identifying other items for which a relatively large portion of users that
20 purchased the popular item also purchased the other item.
4. The method of Claim 3, wherein step (a) comprises calculating, for each (popular item, other item) pair, a similarity score which reflects a number of users that purchased both the popular item and the other item relative to a number of users that purchased at least one of the popular item and the other item.
- 25 5. The method of Claim 1, wherein step (a) comprises calculating, for each of a plurality of item pairs (item_A, item_B), a similarity score which reflects a number of users that indicated an interest in both item_A and item_B relative to a number of users that indicated an interest in at least one of item_A and item_B.
- 30 6. The method of Claim 5, wherein step (a) comprises storing the similarity scores within the data structure, and step (d) comprises using the similarity scores to combine sets of similar items.
7. The method of Claim 1, wherein step (a) is repeated periodically, so that item similarities reflected in the data structure reflect current preferences of the community of users.

8. The method of Claim 1, wherein steps (b)-(e) are performed substantially in real-time in response to a request by the user.

9. The method of Claim 1, wherein the computer system implements online shopping carts for allowing users to select and hold items for possible purchase, and step (b) 5 comprises selecting items that are currently in the user's shopping cart.

10. The method of Claim 1, wherein step (b) comprises identifying a plurality of items that were previously purchased by the user.

11. The method of Claim 1, wherein step (b) comprises selecting only items that have been placed within a designated shopping cart of a plurality of shopping carts of the user, 10 the method thereby generating recommendations that are specific to a role of the designated shopping cart.

12. The method of Claim 1, wherein step (d) comprises weighting at least some of the similar items sets identified in step (c) based on information about the user's affinity for corresponding items of known interest.

15 13. The method of Claim 1, wherein the computer system implements online shopping carts for allowing users to select and hold items for possible purchase, and keeps track of deletions of items from shopping carts, and wherein step (e) further comprises selecting to recommend to the user an item that was previously removed from the user's shopping cart.

20 14. The method of Claim 1, wherein step (a)-(e) are performed without requiring any users to have rated items.

15. The method of Claim 1, further comprising filtering out similar items identified in step (c) to remove items already purchased by the user.

25 16. The method of Claim 1, further comprising filtering out similar items identified in step (c) to remove items that do not fall within an item category designated by the user.

17. In a computer system that provides access to a database of items, a system for recommending items to users, comprising:

30 a first process which determines similarities between items by at least analyzing historical data that reflects item interests of a community of users, the first process generating a data structure which maps items to sets of similar items; and

a second process which generates personal recommendations for a user by accessing the data structure to identify similar items sets that correspond to items

known to be of interest to the user, and by combining the identified similar items sets to generate a list of recommended items.

18. The system of Claim 17, wherein the first process determines a similarity between a pair of items, item_A and item_B, by at least calculating a similarity score which
5 reflects a number of users that are interested in both item_A and item_B relative to a number of users the are interested in item_A or item_B.

19. The system of Claim 18, wherein the first process determines a similarity between a pair of items, item_A and item_B, by at least calculating a score value which reflects a number of users that purchased both item_A and item_B relative to a number of
10 users that purchased item_A or item_B.

20. The system of Claim 17, wherein the first process generates and stores within the data structure similarity scores that indicate degrees of similarity between items, and the second process uses the similarity scores to combine sets of similar items.

21. The system of Claim 17, wherein the first process is an off-line process which
15 executes separately from the second process, and the second process generates recommendations substantially in real-time in response to requests from users.

22. The system of Claim 17, wherein the first process is executed periodically to generate a new data structure, so that item similarities reflected in the data structure reflect current preferences of the community of users.

20 23. The system of Claim 17, wherein the computer system implements online shopping carts for allowing users to select and hold items for possible purchase, and wherein the second process generates recommendations for the user based on items that are currently in the user's shopping cart.

24. The system of Claim 17, wherein the second process weights at least some of
25 the identified similar items sets based on information about the user's affinity for corresponding items of known interest.

25. The system of Claim 17, wherein the computer system allows a user to create multiple shopping carts within a single account, and the second process generates shopping cart specific recommendations to allow a user with multiple shopping carts to obtain
30 recommendations specific to a role of a particular shopping cart.

26. The system of Claim 17, wherein the system generates personal recommendations without requiring users to rate items.

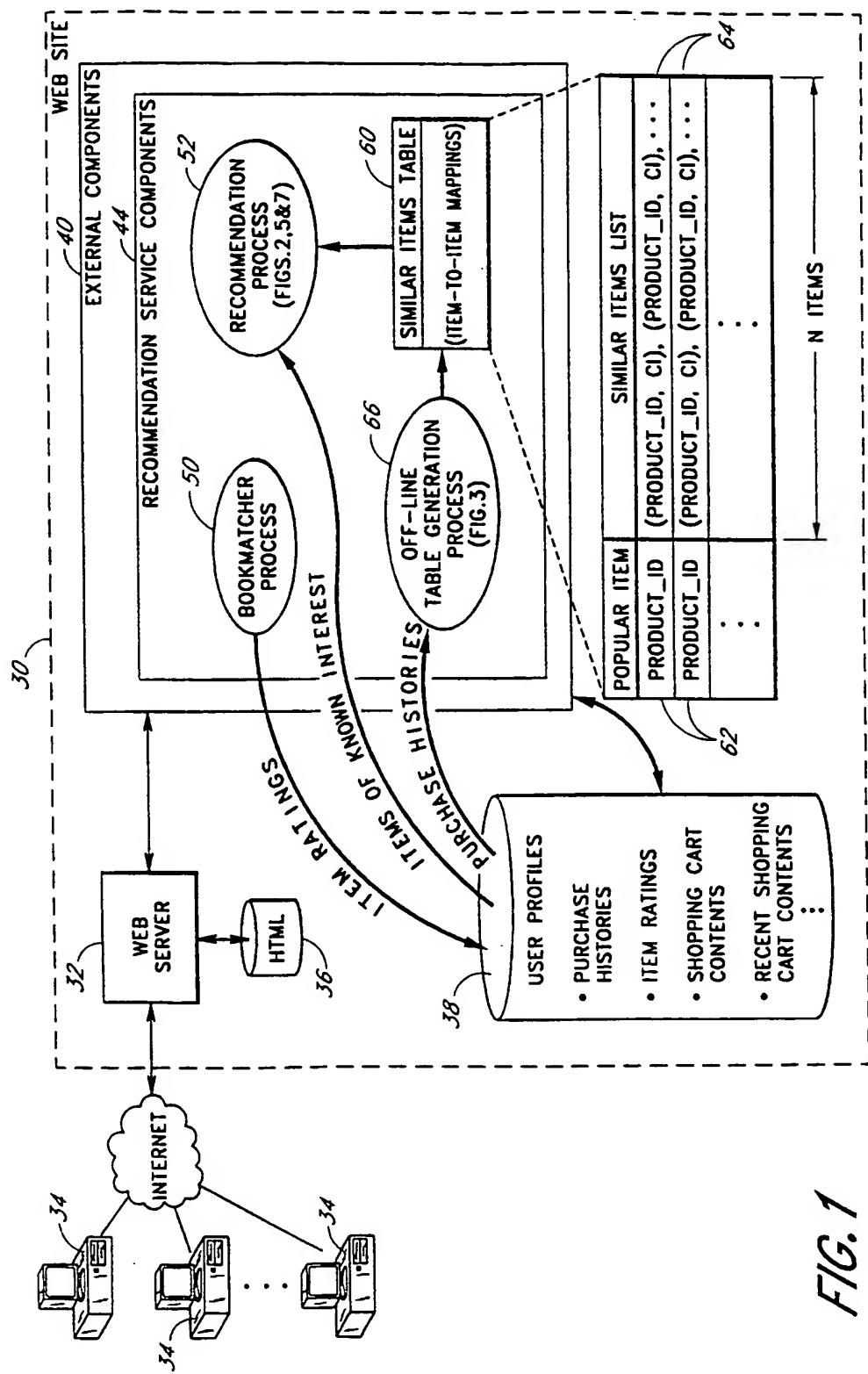
27. The system of Claim 17, wherein the second process filters out items already purchased by the user from the similar items sets identified from the data structure.

28. The system of Claim 17, wherein the second process filters out items from the similar items sets identified from the data structure based on item categories specified by users.

29. A computer-implemented method of identifying items to recommend to a user
5 from a database of items that are available for purchase, comprising:

for each of a plurality of items that are known to be of interest to the user,
identifying a set of items that are deemed to be similar to the respective item based at least upon a collective analysis of purchase histories of a plurality of users; and
combining the resulting plurality of sets.

1 / 7



2/7

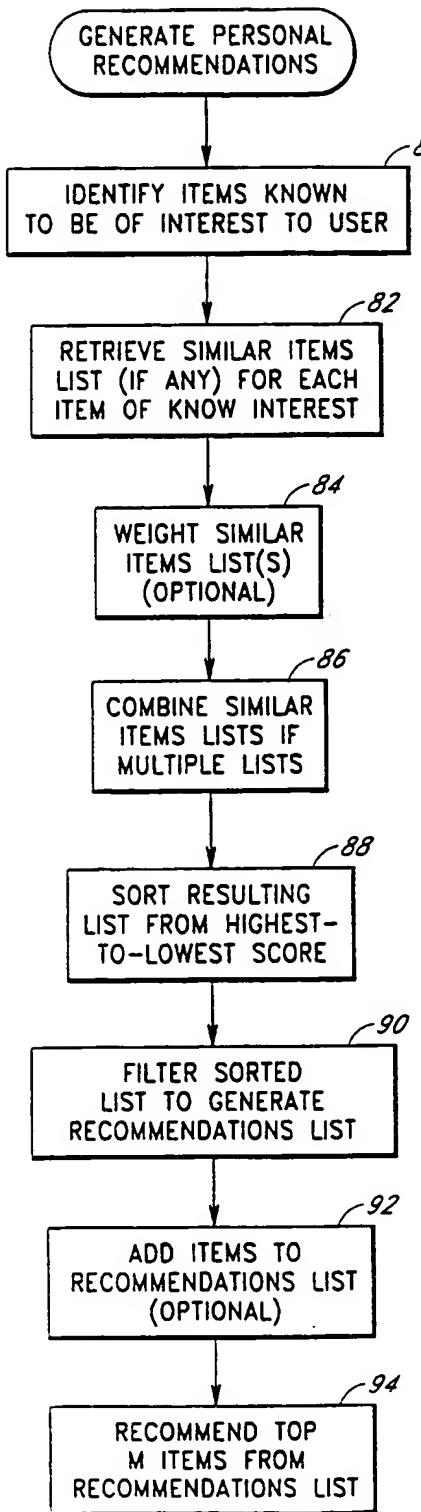
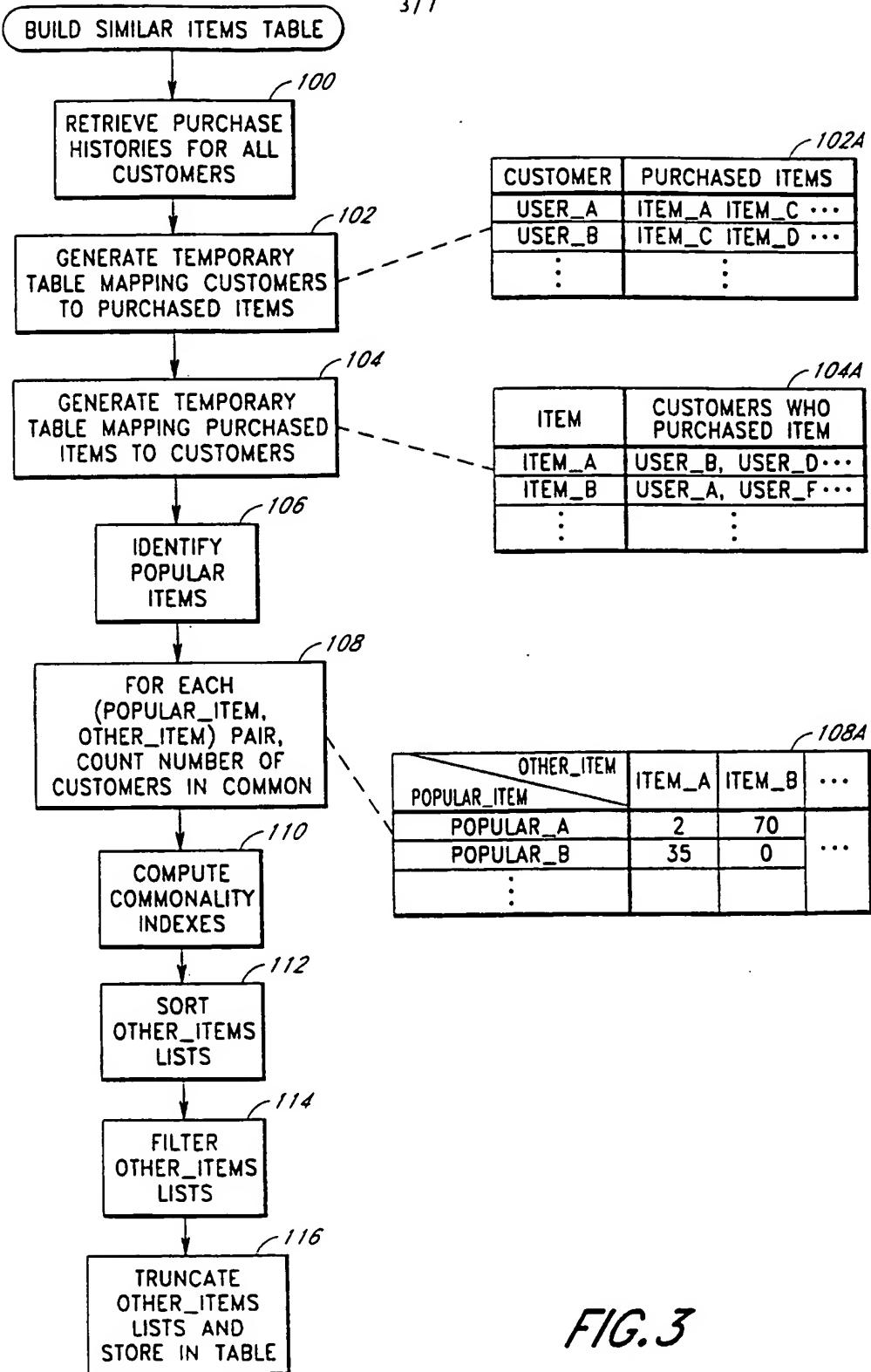


FIG.2

3/7



4 / 7

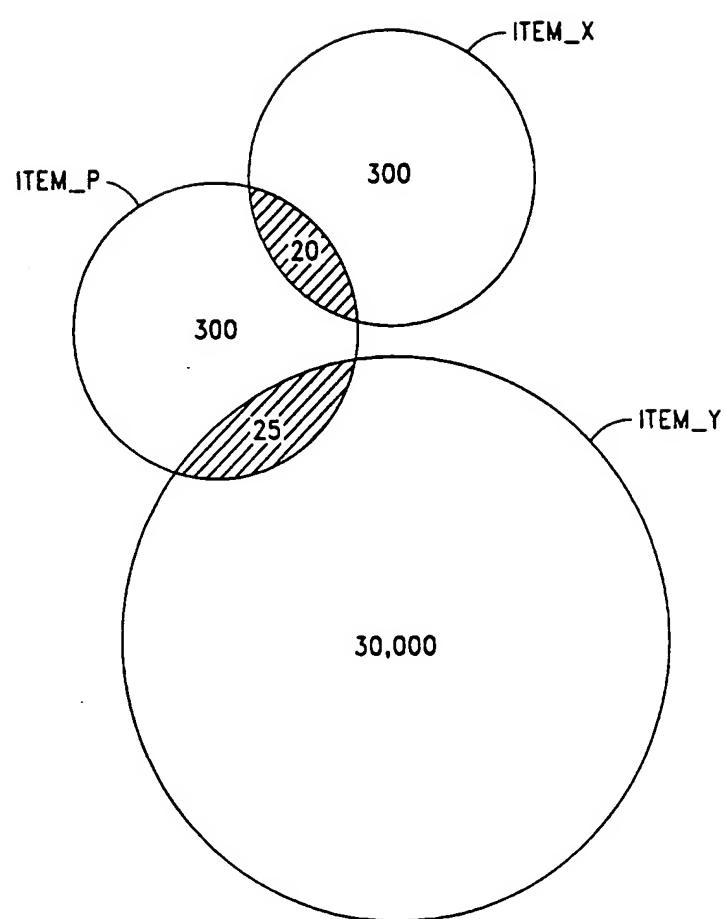


FIG. 4

5/7

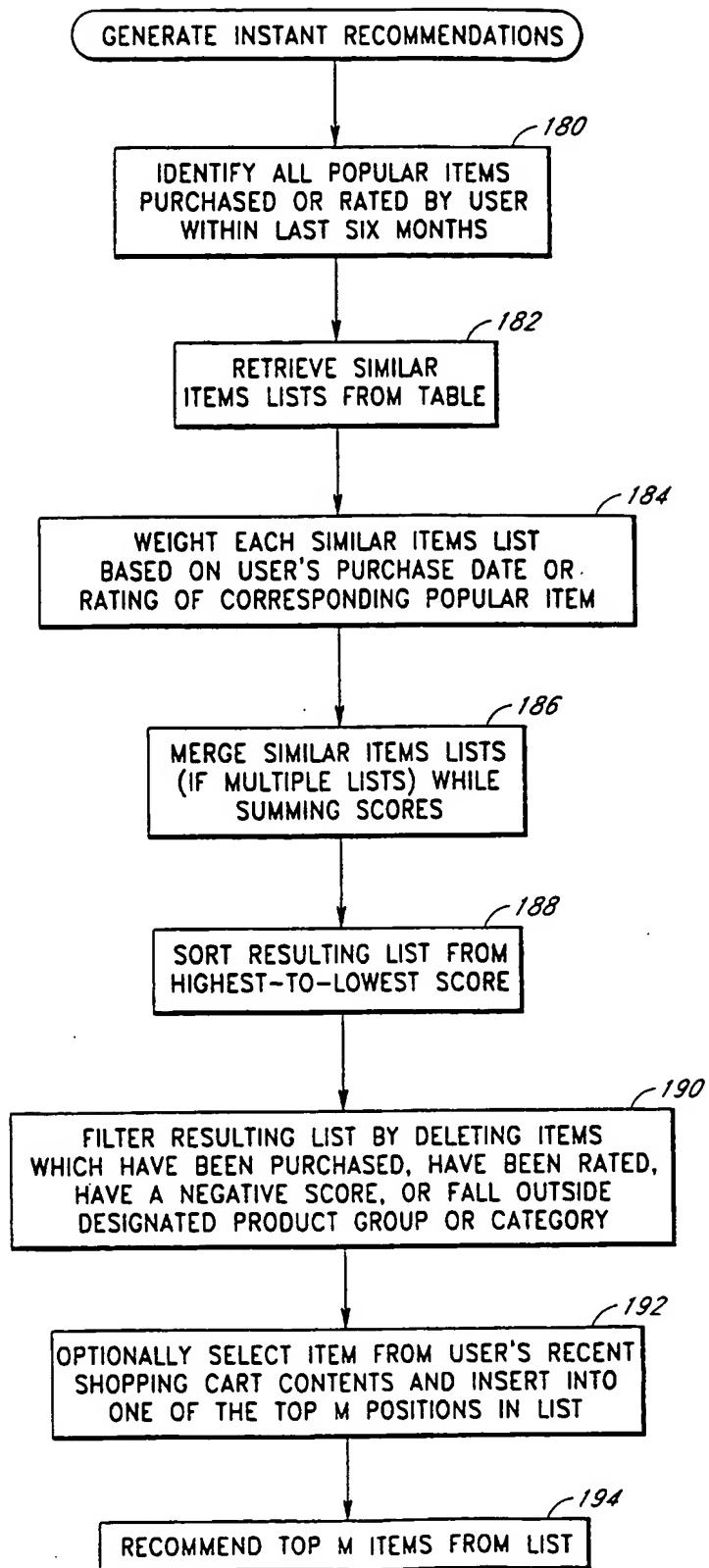


FIG. 5

6 / 7

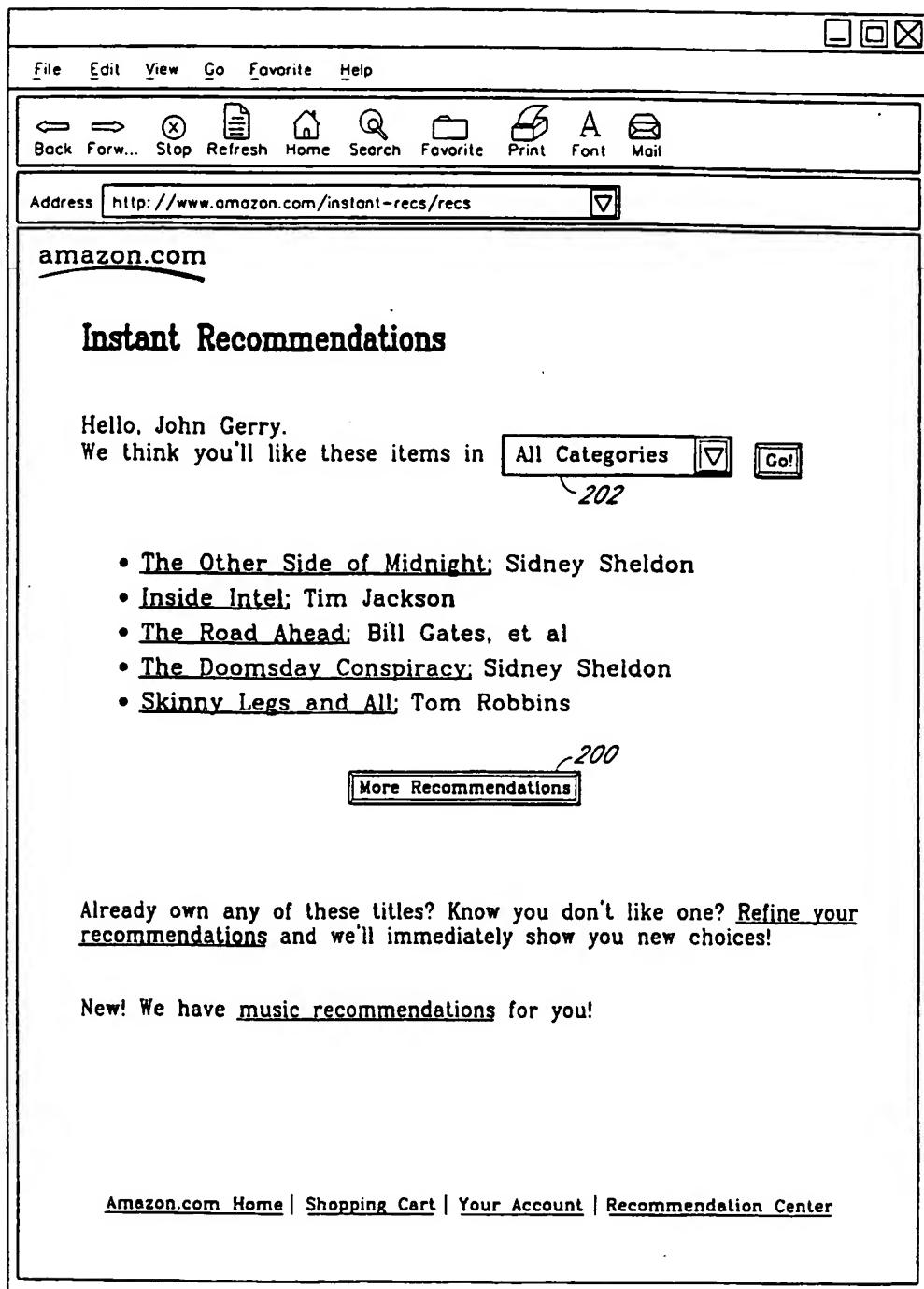


FIG. 6

7/7

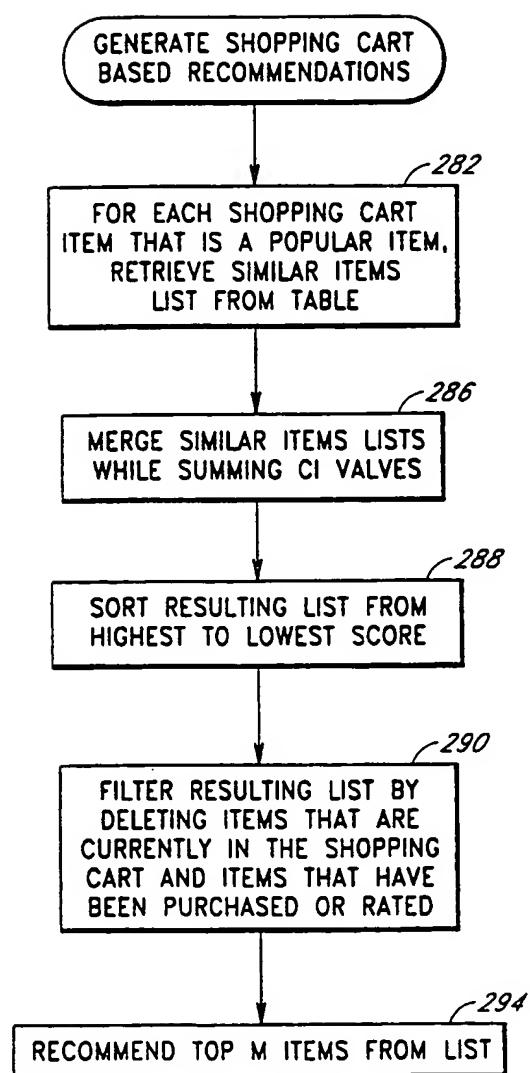


FIG. 7

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US99/20974

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : G06F 17/60
 US CL : 235/383; 380/24; 455/S.1, 450; 705/14, 26, 27; 707/3, 102; 709/227

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
 U.S. : 235/383; 380/24; 455/S.1, 450; 705/14, 26, 27; 707/3, 102; 709/227

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
 APS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 5,583,763 A (ATCHESON et al.) 10 December 1996 (10.12.96), see entire document	1-30
Y	US 5,749,081 A (WHITEIS) 05 May 1998 (05.05.98), see entire document.	1-30
Y	US 5,774,670 A (MONTULLI) 30 June 1998 (30.06.98), see entire document.	11, 25
Y	US 5,790,426 A (ROBINSON) 04 August 1998 (04.08.98), see entire document.	3-6, 18-19
Y, P	US 5,905,973 A (YONEZAWA et al.) 18 May 1999 (18.05.99), see entire document.	9, 13, 23
Y, P	US 5,909,023 A (ONO et al.) 01 June 1999 (01.06.99), see entire document.	2, 10
Y	EP 0 827 063 A1 (JORNA et al.) 04 March 1998 (04.03.98), see entire document.	9, 11, 13, 23, 25
A	US 4,870,579 A (HEY) 26 September 1989 (26.09.89), see entire document.	1-30
A	US 4,992,940 A (DWORKIN) 12 February 1991 (12.02.91), see entire document.	1-30
A	US 4,996,642 A (HEY) 26 February 1991 (26.02.91), see entire document.	1-30
A, P	US 5,884,282 A (ROBINSON) 16 March 1999 (16.03.99), see entire document.	1-30
A, P	US 5,909,492 A (PAYNE et al.) 01 June 1999 (01.06.99), see entire document.	1-30
A, P	US 5,918,014 A (ROBINSON) 29 June 1999 (29.06.99), see entire document.	1-30
A	EP 0 751 471 A1 (LASHKARI et al.) 02 January 1997 (02.01.97), see entire document.	1-30
A	DELGADO, J. Intelligent Collaborative Information Retrieval.	1-30
A	DELGADO, J. Content-Based Collaborative Information Filtering: Actively Learning to Classify and Recommend Documents.	1-30
A	BALABANOVIC, M. and SHOHAM, Y. Content-Based, Collaborative Recommendation, Communications of the ACM, Vol. 40, No. 3, pages 66-72.	1-30

<input type="checkbox"/>	Further documents are listed in the continuation of Box C.	<input type="checkbox"/>	See patent family annex.
*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier application or patent published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 22 November 1999 (22.11.1999)	Date of mailing of the international search report 22 DEC 1999
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703)305-3230	Authorized officer Todd Voeltz <i>James R. Matthews</i> Telephone No. (703) 305-9714